DTU Library

The Relationship Between Background Noise Envelope Power and Speech Intelligibility in Adverse Conditions

Jørgensen, Søren; Decorsiere, Remi Julien Blaise; MacDonald, Ewen; Dau, Torsten

Publication date:
2013

Citation (APA):
Jørgensen, S., Decorsiere, R. J. B., MacDonald, E., & Dau, T. (2013). *The Relationship Between Background Noise Envelope Power and Speech Intelligibility in Adverse Conditions*. Poster session presented at 36th Annual Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, United States.

The Relationship Between Background Noise Envelope Power and Speech Intelligibility in Adverse Conditions

Poster #216

Presented at the 36th Annual Midwinter Meeting of the Association for Research in Otolaryngology

February 16, 2013 Baltimore Marriott Waterfront Hotel Baltimore, Maryland, USA

Authors:
Søren Jørgensen, Rémi Decorsière, Ewen MacDonald, Torsten Dau
Centre for Applied Hearing Research, Technical University of Denmark

Background
Classical models of speech perception, such as in the speech intelligibility index (SII), assume that the long-term speech-to-noise ratio and the spectral characteristics of the background noise determine intelligibility. In addition, the integrity of speech envelope fluctuations has been known to play a role in conditions with reverberation. However, recent studies have demonstrated that the inherent envelope fluctuations of the background noise also play a crucial role for the speech intelligibility in adverse conditions (Jørgensen and Dau, JASA 2011; Stone et al., JASA 2011, 2012). This is reflected in the speech-based envelope power spectrum model (sEPSM) proposed by Jørgensen and Dau (JASA, 2011), which uses the signal-to-noise envelope power ratio (SNR_{env}) as the main predictor of speech intelligibility. Here, a multi-resolution sEPSM (mr-sEPSM) is presented and evaluated in conditions with stationary and fluctuating interferers as well as in conditions where the background noise envelope power is varied systematically using a modulation processing tool (Decorsière et al., submitted). The central hypothesis is that the speech intelligibility is a monotonic function of SNR_{env}.

Methods
Model predictions were compared to literature and own data for speech mixed with three stationary and five fluctuating interferers. In addition, predictions were compared to new data obtained for speech mixed with a modified stationary speech-shaped noise (SSN) where the noise envelope power was attenuated or amplified in the modulation-frequency range between 4 and 20 Hz, while keeping the overall energy of the signal the same.

Results
Lower speech reception thresholds (SRT), i.e., better speech intelligibility, were obtained with the fluctuating noises in comparison to the stationary SSN, demonstrating speech masking release (MR). When the SSN envelope power was attenuated by 20 dB, the SRT decreased by 2 dB relative to the unprocessed SSN, suggesting a release from modulation masking. When the SSN envelope power was amplified by 20 dB, the noise took on a fluctuating character and led to an MR of 10 dB. The mr-sEPSM accounted well for the data in all conditions.

Conclusion
The envelope power of the noise alone plays a crucial role for speech intelligibility in adverse conditions. The mr-sEPSM predictions were in good agreement with the data obtained in the various experimental conditions with stationary and non-stationary interferers, supporting the hypothesis that the SNR_{env} is a powerful objective metric for speech intelligibility prediction.

#216

36th MIDWINTER MEETING

BALTIMORE, MARYLAND

THE RELATIONSHIP BETWEEN BACKGROUND NOISE ENVELOPE POWER AND SPEECH INTELLIGIBILITY IN ADVERSE CONDITIONS

Søren Jørgensen, Rémi Decorsière, Ewen MacDonald, and Torsten Dau

Technical University of Denmark
Centre for Applied Hearing Research
Department of Electrical Engineering

contact: sjor@elektro.dtu.dk

INTRODUCTION

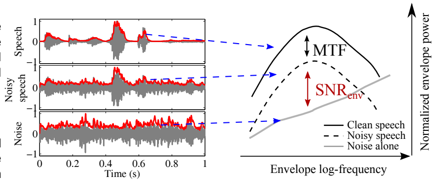
The standardized speech intelligibility index (SII, ANSI S3.5-1997) assumes that the speech-to-noise ratio and effects of spectral masking determine speech intelligibility in conditions with a stationary background noise, while the integrity of speech envelope fluctuations plays a role in conditions with reverberation (Houtgast and Steeneken, 1985). In addition, temporal fluctuations of the background noise may lead to speech "masking release" (Miller and Licklider, 1950), which has been accounted for by the extended (short-term) SII (ESII, Rhebergen and Versfeld, 2005). However, recent studies have suggested that the inherent envelope fluctuations of a seemingly stationary noise may lead to "modulation masking", affecting speech intelligibility in adverse conditions (e.g., Jørgensen and Dau, 2011; Stone et al., 2012). Modulation masking is reflected in the signal-to-noise envelope power ratio (SNR_{env}), which is included in the speech-based envelope power spectrum model (sEPSM) proposed by Jørgensen and Dau (2011). Here, a multi-resolution version of the sEPSM (mr-sEPSM) is presented and evaluated in conditions with stationary and fluctuating interferers as well as in conditions where the envelope fluctuations of a stationary noise are attenuated or enhanced systematically using an envelope processing tool.

BACKGROUND

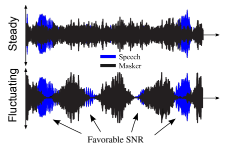
The SNR_{env} is a measure of the relative strength of the speech and the noise envelope fluctuations, estimated from the normalized envelope power P_{env} of the noisy speech and the noise alone:

$$SNR_{env} = \frac{P_{env,S+N} - P_{env,N}}{P_{env,N}}$$

In comparison, the MTF (e.g., Houtgast and Steeneken, 1985) measures the difference between the clean speech and noisy speech envelopes.



Speech masking release



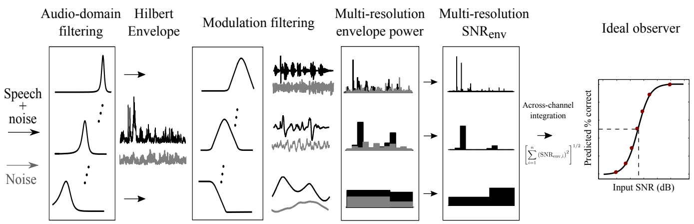
Speech masking release refers to the better intelligibility obtained in conditions with a fluctuating noise compared to a stationary speech-shaped noise (SSN). It has classically been explained by the listener's ability to "glimpse" (Cooke, 2006) the speech when there is a favorable SNR. However, recent work by Stone et al. (2012) suggests that speech masking release actually represents a release from *modulation masking*, since it mainly occurs when the un-modulated masker contains inherent envelope fluctuations.

HYPOTHESIS

It is hypothesized that the noise envelope fluctuations play an important role for the intelligibility of speech in stationary and fluctuating noise, and that there is a monotonic relationship between SNR_{env} and speech intelligibility in noise. It is argued that an attenuation of the noise envelope fluctuations should lead to a release from modulation masking, which cannot be accounted for by the concept of glimpsing.

MODEL

Multi-resolution speech-based envelope power spectrum model (mr-sEPSM)



Noisy speech (black) and noise alone (gray) are processed separately through a filterbank of 22 gammatone filters (f_c = 63-8000 Hz) followed by envelope extraction. Sub-band envelopes are processed by a modulation filterbank (9 filters with f_{mod} = 1-256 Hz). The envelope power is computed at the output of each modulation filter with a temporal-window duration of $1/f_{mod}$. The multi-resolution SNR_{env} is computed, averaged across time, and integrated across modulation filters and audio-domain filters. Finally, the overall SNR_{env} is converted to a percent-correct prediction assuming an "ideal observer".

CONDITIONS

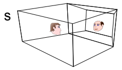
Stationary interferers: **SSN** Speech-shaped noise, **Car**-noise, **Bottle**-noise.

Fluctuating interferers: **Cafe** noise, **SAM** - sinusoudal amplitude modulated noise, **SMN** - speech modulated noise, **RT** - time-reversed interfering talker, **ISTS** - International speech test signal.

Spectral subtraction: α = 0, 0.5, 1, 2, 4, 8.

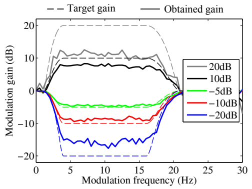
$$\hat{S}(f) = \left[P_{S+N}(f) - \alpha \hat{P}_N(f)\right]^{1/2}$$

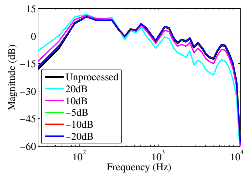
Reverberation: T_{30} = 0, 0.4, 0.7, 1.3, 2.3 s



Reverberent impulse responses simulated with ODEON

Envelope filtering: SSN was processed to have target modulation gains of -20, -10, -5, 10, and 20 dB within the modulation-frequency range between 4 and 16 Hz, using the method described in Decorsière et al. (2013)

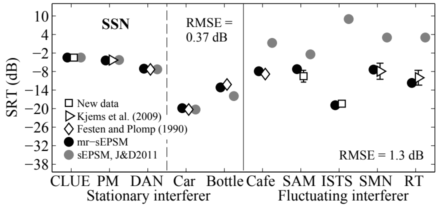




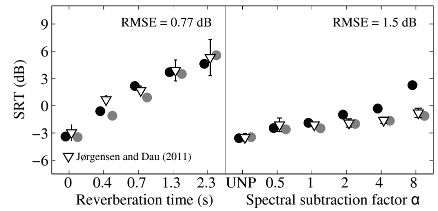
Left: target and obtained modulation transfer functions of the processed noise. The obtained gain deviate from the target for large gains.

Right: long-term spectra of the processed noise stimuli. The spectra are identical for negative gains but deviate slightly from the unprocessed noise for positive gains.

RESULTS

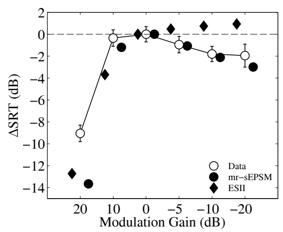


Measured speech reception thresholds (SRT) (open symbols) and predictions from the mr-sEPSM and the original "long-term" sEPSM (filled black and gray circles). The mr-sEPSM accounts well for both stationary and fluctuating interferers, demonstrating an advantage over the original sEPSM.



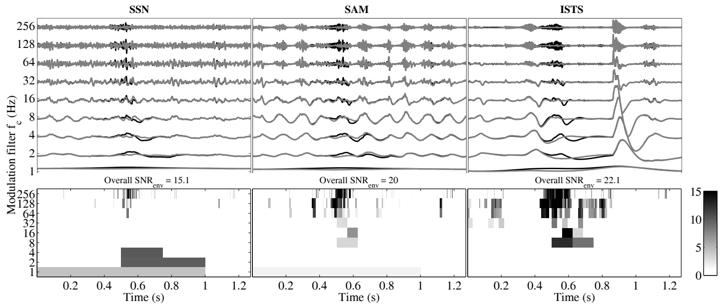
Measured SRTs (open symbols) and predictions from the mr-sEPSM and the original "long-term" sEPSM (filled black and gray circles). Both model versions account well for the conditions.

Envelope filtering

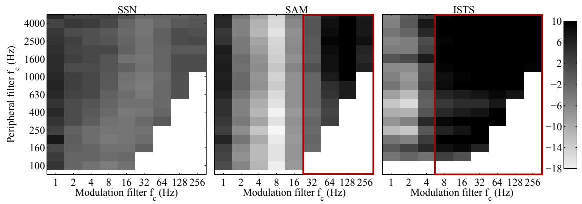


Measured change in SRT (ΔSRT) from the unprocessed condition (open symbols) and predictions from the mr-sEPSM (filled black circles). A target amplification of 20 dB led to an improvement in SRT of 9 dB, while an attenuation of 20 dB led to a small but significant improvement of 2 dB. The mr-sEPSM accounts for these effects, while predictions from the ESII (black diamonds), which is based on glimpsing, fails to account for the conditions with attenuated noise fluctuations (negative gains).

MODEL ANALYSIS



Top panels: Output from the nine modulation filters as a function of time in response to the interferer alone (gray) and the interferer plus speech (black), for three different interferers: SSN (left), SAM (middle), and ISTS (right). Bottom panels: SNR_{env}-values (in dB) per time segment, calculated from the modulation-filter outputs shown in the upper panels. The overall time-averaged SNR_{env} (indicated directly above each of the bottom panels) is higher for the fluctuating noises compared to SSN.



Time-averaged SNR_{env} (in dB), at an input SNR of -5 dB, as a function of peripheral and modulation filters for the SSN (left), SAM (middle), and ISTS (right). Modulation f_c above 4-16 Hz provide the main contribution to predicted intelligibility in fluctuating interferers (red rectangles).

SUMMARY AND DISCUSSION

- Lower SRTs, i.e. better speech intelligibility, were obtained with the fluctuating interferers in comparison to the stationary SSN, demonstrating a speech masking release.

- The conditions with envelope filtering demonstrated that when the SSN was processed to have a target modulation gain of -20 dB, the SRT decreased by 2 dB relative to the unprocessed SSN. This small but significant effect could not be explained by "glimpsing", suggesting a release from modulation masking. When the target modulation gain was 20 dB, the noise exhibited a fluctuating character and led to a masking release of 9 dB.

- The mr-sEPSM accounted well for the data in conditions with stationary and non-stationary interferers, as well as the distorted noisy speech. Moreover, the model accounted for the effects of envelope filtering, in contrast to the ESII, suggesting that the SNR_{env} is a powerful metric for speech intelligibility prediction.

- A model analysis revealed that the key component allowing the mr-sEPSM to account for the fluctuating interferers is the multi-resolution estimation of the SNR_{env}, whereby the short segments in the modulation filters tuned to modulation frequencies above about 16 Hz contributed most to masking release. This suggests, that "glimpsing" can be expained by release from modulation masking in the dips of the interferer, consistent with Stone et al. (2012).

References

American National Standards Inst. (**1997**). ANSI-S3.5: Methods for the calculation of the speech intelligibility index (American National Standards Institute, New York).

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise", J. Acoust. Soc. Am. **119**, 1562–1573.

Decorsière, R., MacDonald, E. N., Søndergaard, P. L., and Dau, T. (**2013**). "Optimization approach to the reconstruction of a signal from a spectrogram-like representation", IEEE Trans. Audio Speech Lang. Process. (submitted).

Festen, J. M. and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing" J. Acoust. Soc. Am., **88**, 1725–1736.

Houtgast, T. and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria", J. Acoust. Soc. Am. **77**, 1069-1077.

Jørgensen, S. and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing" J. Acoust. Soc. Am. **130**, 1475–1487.

Miller, G. A. and Licklider, J. C. R., (**1950**). "The intelligibility of interrupted speech", J. Acoust. Soc. Am. **22**, 167–173.

Kjems, U., Boldt, J.B., Pedersen, M.S., Lunner, T., Wang, D., (**2009**). "Role of mask pattern in intelligibility of ideal binary-masked noisy speech", J. Acoust. Soc. Am. **126**, 1415–1426.

Rhebergen, K. S. and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners", J. Acoust. Soc. Am. **117**, 2181–2192.

Stone, M.A., Füllgrabe, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech", J. Acoust. Soc. Am. **132**, 317-326.